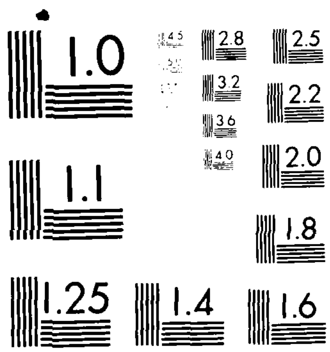
AD-A162 796 EFFECTS OF ESTIMATED NOISE COVARIANCE MATRIX IN OPTIMAL SIGNAL DETECTION (U) PITTSBURGH UNIV PA CENTER FOR MULTIVARIATE ANALYSIS C G KHATRI ET AL. OCT 85 1/1

UNCLASSIFIED TR-85-38 AFOSR-TR-85-1130 F49620-85-C-0008 F/G 9/4 NL


1.0
2.8
2.5
3.2
2.2
3.6
4.0
2.0
1.1
1.8
1.25
1.4
1.6

AD-A162 796

EFFECTS OF ESTIMATED NOISE
COVARIANCE MATRIX IN OPTIMAL
SIGNAL DETECTION

by

C.G. Khatri
Gujarat University, Ahmedabad

and

C. Radhakrishna Rao
University of Pittsburgh
Pittsburgh, PA 15260

Center for Multivariate Analysis

University of Pittsburgh

85 12 30 0 . .

EFFECTS OF ESTIMATED NOISE COVARIANCE MATRIX IN OPTIMAL SIGNAL DETECTION

by

C.G. Khatri
Gujarat University, Ahmedabad

and

C. Radhakrishna Rao
University of Pittsburgh
Pittsburgh, PA 15260

October 1985

Technical Report 85-38

Center for Multivariate Analysis
515 Thackeray Hall
University of Pittsburgh
Pittsburgh, PA 15260

*This work is supported by Contract N00014-85-K-0292 of the Office of Naval Research and Contract F49620-85-C-0008 of the Air Force Office of Scientific Research. The United States Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright notation hereon.

EFFECTS OF ESTIMATED NOISE COVARIANCE MATRIX IN OPTIMAL SIGNAL DETECTION

by

C.G. Khatri
Gujarat University, Ahmedabad

and

C. Radhakrishna Rao
University of Pittsburgh
Pittsburgh, PA 15260

ABSTRACT

There is loss of efficiency when an estimated noise covariance matrix is used in the place of the unknown true noise covariance matrix in the construction of the optimum filter for signal detection. In the case of detecting a single signal specified by a real or a complex vector, we investigate the extent of this loss by obtaining an exact confidence bound for the realized signal to noise ratio. We also give an estimate of this ratio which is useful in optimum selection of features. Some of these results are extended to the case of discrimination between a number of given signals.

1. INTRODUCTION

Reed, Mallet and Brennan (1974) studied the loss of power in signal detection when the noise covariance matrix is unknown and the estimated matrix from sampled data on noise is used in the construction of the optimum filter or the linear discriminant function. This was done by computing the expected value of the signal to noise ratio based on the estimated filter and comparing it with the corresponding ratio when the covariance matrix is known. In this paper, we extend the study of the above authors in several directions.

An exact confidence bound is provided for the realized signal to noise ratio when an estimated filter is used. A test is given for examining whether a given set of features is sufficient for signal detection. A criterion is provided for optimum selection of features. Finally, the problem of discrimination with multiple alternative signals is discussed. We consider both the cases where the signal is represented by a real or a complex vector.

The following notations are used. A' denotes the transpose of a matrix A when its elements are real and A* the conjugate transpose of A when its elements are complex.

i) $X \sim N_p(\mu,\Sigma)$, i.e., a real p-vector X has a p-variate real normal distribution with the probability density function (p.d.f.)

$$(2\pi)^{-p/2}|\Sigma|^{-\frac{1}{2}}\exp\left[-\tfrac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right]. \tag{1.1}$$

ii) $X \sim \tilde{N}_p(\mu,\Sigma)$, i.e., a complex vector X has a p-variate complex normal distribution with the p.d.f.

$$(\pi)^{-p}|\Sigma|^{-1}\exp\left[-(x-\mu)^*\Sigma^{-1}(x-\mu)\right]. \tag{1.2}$$

iii) $Y \sim N_{r,s}(M,\Sigma,V)$, i.e., a real r x s matrix Y has the p.d.f.

$$(2\Pi)^{-rs/2}|\Sigma|^{-s/2}|V|^{-r/2}\exp[-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}(Y-M)V^{-1}(Y-M)'] . \tag{1.3}$$

iv) $Y \sim \tilde{N}_{r,s}(M,\Sigma,V)$, i.e., a complex r x s matrix has the p.d.f.

$$(\Pi)^{-rs}|\Sigma|^{-s}|V|^{-r}\exp[-\mathrm{tr}\Sigma^{-1}(Y-M)V^{-1}(Y-M)^*] . \tag{1.4}$$

v) $S \sim W_p(f,\Sigma)$, i.e., a real p x p positive definite matrix S has the Wishart distribution on f degrees of freedom with the p.d.f.

$$2^{-pf/2}[\Gamma_p(f/2)]^{-1}|\Sigma|^{-f/2}|S|^{(f-p-1)/2}\exp(-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}S) \tag{1.5}$$

where

$$\Gamma_p(a) = \Pi^{p(p-1)/4}\prod_{i=1}^{p}(a-\frac{i-1}{2}) .$$

vi) $S \sim \tilde{W}_p(f,\Sigma)$, i.e., a complex p x p positive definite matrix S has the complex Wishart distribution with the p.d.f.

$$[\tilde{\Gamma}_p(f)]^{-1}|\Sigma|^{-f}|V|^{f-p}\exp(-\mathrm{tr}\Sigma^{-1}S) \tag{1.6}$$

where

$$\tilde{\Gamma}_p(a) = \Pi^{p(p-1)/2}\prod_{i=1}^{p}(a-i+1) .$$

vii) $S \sim W_p^g(f,\Sigma)$, i.e., a real p x p positive definite matrix has the p.d.f.

$$|\Sigma|^{-f/2}|S|^{(f-p-2)/2}g(-\tfrac{1}{2}\mathrm{tr}\Sigma^{-1}S) . \tag{1.7}$$

viii) $S \sim \tilde{W}_p^g(f,\Sigma)$, i.e., a complex p x p positive definite matrix S has the p.d.f.

$$|\Sigma|^{-f}|S|^{f-p}g(-\mathrm{tr}\Sigma^{-1}S) . \tag{1.8}$$

2. SOME MULTIVARIATE DISTRIBUTIONS

In this section we derive some new multivariate distributions which arise in the study of problems of signal detection. The actual applications are discussed in Section 3.

Consider the p x p positive definite (p.d.) matrices

$$\begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix}, \qquad \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix} \tag{2.1}$$

partitioned by the first r and the rest s = p - r of rows and columns, the Schur complements of order r x r

$$S_{1.2} = S_{11} - S_{12}S_{22}^{-1}S_{21}, \ \Sigma_{1.2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21} \tag{2.2}$$

and the regression coefficients of order r x s

$$b = S_{12}S_{22}^{-1}, \quad \beta = \Sigma_{12}\Sigma_{22}^{-1}. \tag{2.3}$$

We have the following lemmas which follow on standard lines (see Rao (1973, pp. 538-539) and Srivastava and Khatri (1979, p. 79)).

Lemma 1. Let $S \sim W_p(f,\Sigma)$ where $p = r + s$ and $S_{ij}, S_{1.2}, \Sigma_{1.2}, b$ and β be as defined in (2.1)-(2.3). Then the following hold:

1.) $S_{1.2}$ and (b,S_{22}) are independently distributed with

$$S_{1.2} \sim W_r(f-s, \Sigma_{1.2}) \tag{2.4}$$

$$S_{22} \sim W_s(f, \Sigma_{22}) \tag{2.5}$$

and the conditional distribution of the r x s matrix b given S_{22} is

$$b \sim N_{r,s}(\beta,\Sigma_{1.2},S_{22}^{-1}). \tag{2.6}$$

2.) The unconditional (marginal) p.d.f. of b obtained by integrating over S_{22} is

$$\frac{\Gamma_s(\frac{f+r}{2})}{\Pi^{rs/2}\Gamma_s(\frac{f}{2})} |\Sigma_{22}|^{-f/2}|\Sigma_{1.2}|^{-s/2}|\Sigma_{22}^{-1}+(b-\beta)'\Sigma_{1.2}^{-1}(b-\beta)|^{-(f+r)/2}$$

which we denote by

$$T_{r,s}(\beta,f,\Sigma_{1.2},\Sigma_{22}). \tag{2.7}$$

If $b_1 = \Sigma_{1.2}^{-\frac{1}{2}}(b-\beta)\Sigma_{22}^{\frac{1}{2}}$, where $\Sigma_{1.2}^{\frac{1}{2}}$ and $\Sigma_{22}^{\frac{1}{2}}$ represent symmetric square roots, then

$$b_1 \sim T_{r,s}(o,f,I_r,I_s). \tag{2.8}$$

3.) If $u = (I_r+b_1b_1')^{-\frac{1}{2}}\, b_1 = b_1(I_s+b_1'b_1)^{-\frac{1}{2}}$, then the Jacobian of the transformation from b_1 to u is $|I_r-UU'|^{-(r+s+1)/2}$ and hence the p.d.f. of u, derived from (2.8), is

$$\frac{\Gamma_s(\frac{f+r}{2})}{\Pi^{rs/2}\Gamma_s(\frac{f}{2})} |I_r-UU'|^{(f-s-1)/2} \tag{2.9}$$

which we denote by

$$U_{r,s}(\frac{f+r}{2}). \tag{2.10}$$

4.) If $s \geq r$, the p.d.f. of $B = (I_r+b_1b_1')^{-1}$, derived on standard lines, is

$$[\beta_r(\frac{f+r-s}{2}, \frac{s}{2})]^{-1}|B|^{(f-s-1)/2}|I-B|^{(s-r-1)/2}$$

where

$$\beta_r(a,b) = \frac{\Gamma_r(a)\Gamma_r(b)}{\Gamma_r(a+b)},$$

which is the r-variate beta distribution denoted by

$$B_r(\frac{f+r-s}{2}, \frac{s}{2}). \tag{2.11}$$

If $b_2 = (b-\beta)\Sigma_{22}^{\frac{1}{2}}$, then the p.d.f. of $B_0 = (\Sigma_{1.2}+b_2b_2')^{-1}$ is

$$[\beta_r(\frac{f+r-s}{2}, \frac{s}{2})]^{-1}|\Sigma_{1.2}|^{(f-1)/2}|B_0|^{(f-s-1)/2}|\Sigma_{1.2}^{-1}-B_0|^{(s-r-1)/2}$$

which will be referred to as

$$B_r(\frac{f+r-s}{2}, \frac{s}{2}; \Sigma_{1.2}^{-1}). \tag{2.12}$$

Lemma 2. If $S \sim \tilde{W}_p(f,\Sigma)$ where $p = r + s$, then $S_{1.2}$ and (b_1,S_{22}) are independently distributed, and the distributions of the various statistics considered in Lemma 1 are as follows.

1.) $$S_{1.2} \sim \tilde{W}_r(f-r,\Sigma_{1.2}),\ S_{22} \sim \tilde{W}_s(f,\Sigma_{22}). \tag{2.13}$$

The conditional distribution of b given S_{22} is

$$b \sim \tilde{N}_{r,s}(\beta,\Sigma_{1.2},S_{22}^{-1}) \tag{2.14}$$

2.) The marginal distribution of b is

$$b \sim \tilde{T}_{r,s}(\beta,f,\Sigma_{1.2},\Sigma_{22})$$

with the p.d.f.

$$\frac{\tilde{\Gamma}_s(f+r)}{\pi^{rs}\tilde{\Gamma}_s(f)}|\Sigma_{22}|^{-f}|\Sigma_{1.2}|^{-s}|\Sigma_{22}^{-1} + (b-\beta)^*\Sigma_{1.2}^{-1}(b-\beta)|^{-(f+r)}. \tag{2.15}$$

$$b_1 = \Sigma_{1.2}^{-\frac{1}{2}}(b-\beta)\Sigma_{22}^{\frac{1}{2}} \sim \tilde{T}_{r,s}(o,f,I_r,I_s). \tag{2.16}$$

3.) If $u = b_1(I_s+b_1^*b_1)^{-\frac{1}{2}} = (I_r+b_1b_1^*)^{-\frac{1}{2}}b_1$, then its p.d.f. is

$$\frac{\tilde{\Gamma}_s(f+r)}{\pi^{rs}\tilde{\Gamma}_s(f)}|I_r-UU^*|^{f-s}$$

which is denoted by

$$u \sim \tilde{U}_{r,s}(f+r). \tag{2.17}$$

4.) If $s \geq r$, the p.d.f. of $B = (I_r+b_1b_1^*)^{-1}$, *derived on standard lines*, is

$$[\tilde{\beta}_r(f+r-s,s)]^{-1}|B|^{f-s}|I-B|^{s-r}$$

which will be referred to as r-variate complex beta distribution

$$\tilde{B}_r(f+r-s,s). \tag{2.18}$$

Writing $b_2 = (b-\beta)\Sigma_{22}^{\frac{1}{2}}$, the p.d.f. of $B_0 = (\Sigma_{1.2}+b_2b_2^*)^{-1}$ obtained by a transformation from (2.15) is

$$[\tilde{\beta}_r(f+r-s,s)]^{-1}|\Sigma_{1.2}|^{f}|B_0|^{f-s}|\Sigma_{1.2}^{-1} - B_0|^{s-r}$$

which will be referred to as

$$\tilde{\beta}_r(f+r-s,s;\ \Sigma_{1.2}^{-1}). \tag{2.19}$$

3. MAIN THEOREMS

In this section, we use the results of Section 2 to derive distributions of some functions of a p x r matrix Δ whose columns represent given signals and $f^{-1}S$ the estimated noise covariance matrix of order p x p. These distributions are used in the next section for drawing inferences on the basis of observed data in signal detection. First we consider the real case and quote the corresponding results for the complex case in the remarks following the theorems.

Theorem 1. Let Δ be a p x r given matrix of rank r ($\leq p/2$) and $S \sim W_p(f,\Sigma)$. Define the r x r matrices

$$S_\Delta = (\Delta' S^{-1}\Delta)^{-1},\ \Sigma_\Delta = (\Delta'\Sigma^{-1}\Delta)^{-1} \tag{3.1}$$

$$B = \Sigma_\Delta^{\frac{1}{2}} S_\Delta^{-1}(\Delta' S^{-1}\Sigma S^{-1}\Delta)^{-1} S_\Delta^{-1}\Sigma_\Delta^{\frac{1}{2}}. \tag{3.2}$$

Then S_Δ and B are independently distributed with

$$S_\Delta \sim W_r(f-p+r,\ \Sigma_\Delta) \tag{3.3}$$

$$B \sim B_r\left(\frac{f+r-s}{2},\ \frac{s}{2}\right) \tag{3.4}$$

where the B_r distribution is as defined in (2.11) and $s = p - r$.

Proof. Let Δ_1 be a p x s matrix of rank s(=p-r) such that $\Delta_0 = (\Delta : \Delta_1)$ is nonsingular and $\Delta_1'\Delta = 0$. Then $\Delta_0' S\Delta_0 \sim W_p(f_1\Delta_0'\Sigma\Delta_0)$. Writing

$$\Delta_0' S\Delta_0 = \begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix},\quad \Delta_0'\Sigma\Delta_0 = \begin{pmatrix} \theta_{11} & \theta_{12} \\ \theta_{21} & \theta_{22} \end{pmatrix}$$

$$V_{1.2} = V_{11} - V_{12}V_{22}^{-1}V_{21},\quad \theta_{1.2} = \theta_{11} - \theta_{12}\theta_{22}^{-1}\theta_{21}$$

$$b_1 = \theta_{1.2}^{-\frac{1}{2}}(V_{12}V_{22}^{-1} - \theta_{12}\theta_{22}^{-1})\theta_{22}^{\frac{1}{2}} \tag{3.5}$$

and using Lemma 1

$$V_{1.2} \sim W_r(f-p+r,\ \theta_{1.2}) \tag{3.6}$$

$$(I+b_1b_1')^{-1} \sim B_r\left(\frac{f+r-s}{2},\ \frac{s}{2}\right). \tag{3.7}$$

Further $V_{1.2}$ and $(I+b_1b_1')^{-1}$ are independently distributed. Now

$$V_{1.2} = \Delta'S\Delta - \Delta'S\Delta_1(\Delta_1'S\Delta)^{-1}\Delta_1'S\Delta$$

$$= \Delta'\Delta S_\Delta \Delta'\Delta$$

$$\theta_{1.2} = \Delta'\Delta\Sigma_\Delta \Sigma'\Delta$$

Then from (3.6), $S_\Delta = (\Delta'\Delta)^{-1}V_{1.2}(\Delta'\Delta)^{-1}$ has the desired distribution (3.3).

Further, using the formula

$$\begin{pmatrix} V_{11} & V_{12} \\ V_{21} & V_{22} \end{pmatrix}^{-1} = \begin{pmatrix} 0 & 0 \\ 0 & V_{22}^{-1} \end{pmatrix} + \begin{pmatrix} I \\ -V_{22}^{-1}\ V_{21} \end{pmatrix} V_{1.2}^{-1}\ (I \ :\ -V_{12}V_{22}^{-1})$$

we find, after some computations, that B as defined in (3.2) is the same as $(I + b_1b_1')^{-1}$ with b_1 as in (3.5). Then (3.7) establishes (3.4). Theorem 1 is proved.

<u>Remark 1</u>. If $S \sim W_p^g(f,\Sigma)$ as defined in (1.7), then S_Δ and B as defined in Theorem 1, (3.1) and (3.2), are independently distributed. Further, B has the same p.d.f. (3.4) as in Theorem 1 independently of g, while the same is not true for S_Δ.

<u>Remark 2</u>. Let Δ be a p x r complex matrix of rank $r(\leq p/2)$ and $S \sim \tilde{W}_p(f,\Sigma)$. Then

$$S_\Delta = (\Delta^* S^{-1}\Delta)^{-1} \sim \tilde{W}_r(f-p+r,\ \Sigma_\Delta) \tag{3.8}$$

and

$$B = \Sigma_\Delta^{\frac{1}{2}}\, S_\Delta^{-1}(\Delta^* S^{-1}\Sigma S^{-1}\Delta)^{-1} S_\Delta^{-1}\Sigma_\Delta^{\frac{1}{2}} \sim \tilde{B}_r(f+r-s,s). \tag{3.9}$$

Further S_Δ and B are independently distributed.

Remark 3. If S is complex and has the distribution $\tilde{W}_p^g(f,\Sigma)$, then S_Δ and B as defined in (3.8,3.9) are independently distributed. Further, the distribution of B is as in (3.9) independently of g, while the same is not true for S_Δ.

Theorem 2. Let B be p x p positive definite matrix such that

$$B \sim B_p\left(\frac{f_1}{2}, \frac{f_2}{2}; A\right), \quad 0 \le B \le A.$$

Consider the partitions

$$A = \begin{pmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{pmatrix}, \quad B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix}$$

where A_{11} and B_{11} are r x r matrices, and the Schur complements $A_{2\cdot 1}$ and $B_{2\cdot 1}$. Then the statistics B_{11}, $B_{2\cdot 1}$ and

$$U = (A_{2\cdot 1} - B_{2\cdot 1})^{-\frac{1}{2}}(B_{21} - A_{21}A_{11}^{-1}B_{11})\,[B_{11}^{-1} + (A_{11} - B_{11})^{-1}]^{\frac{1}{2}} \tag{3.10}$$

are independently distributed. Further

$$B_{11} \sim B_r\left(\frac{f_1}{2}, \frac{f_2}{2}; A_{11}\right), \quad 0 \le B_{11} \le A_{11}, \tag{3.11}$$

$$B_{1\cdot 2} \sim B_{p-r}\left(\frac{f_1-r}{2}, \frac{f_2}{2}; A_{2\cdot 1}\right), \quad 0 \le B_{2\cdot 1} \le A_{2\cdot 1}, \tag{3.12}$$

$$U \sim U_{p-r,r}\left(\frac{f_2}{2}\right) \text{ as in (2.10)}. \tag{3.13}$$

The results of Theorem 2 were established by Khatri and Pillai (1965) when $A = I_p$. Their proof can be easily extended to our case by noting that

$$|A| = |A_{11}|\ |A_{2\cdot 1}|,\ |B| = |B_{11}|\ |B_{2\cdot 1}|$$

$$|A-B| = |A_{11}-B_{11}|\ |A_{2\cdot 1}-B_{2\cdot 1}|\ |I_{p-r}-UU'|$$

and then computing the necessary Jacobians of the transformations.

Remark 4. In the complex case, let B and A-B be Hermitian positive definite matrices such that

$$B \sim \tilde{B}_p(f_1,f_2;\ A).$$

Then, B_{11}, $B_{2\cdot 1}$ and U as defined in Theorem 2 are independently distributed. Further

$$B_{11} \sim \tilde{B}_r(f_1,f_2;\ A_{11}), \tag{3.14}$$

$$B_{2\cdot 1} \sim \tilde{B}_{p-r}(f_1-r,f_2;\ A_{2\cdot 1}), \tag{3.15}$$

and

$$U \sim \tilde{U}_{p-r,r}(f_2) \text{ as in (2.17)}. \tag{3.16}$$

Theorem 3. Let X and Y be independent univariate gamma, $G(1,m)$, and beta, $B_1(m-c+1,a)$, variables with the p.d.f.'s

$$\frac{1}{\Gamma(m)} e^{-x}x^{m-1},\ x > 0,\ m > 0 \tag{3.17}$$

and

$$\frac{1}{\beta(m-c+1,a)} \; y^{m-c}(1-y)^{a-1}, \; 0 < y < 1, \; a > 0, \; m-c+1 > 0. \tag{3.18}$$

Then the p.d.f. of Z = XY is

$$\frac{e^{-z}z^{m-1}}{\Gamma(m)} \; \frac{\Gamma(a+m-c+1)}{\Gamma(m-c+1)} \; \Psi(a,c;\, z) \tag{3.19}$$

where Ψ is the confluent hypergeometric function of the second kind defined by

$$\Psi(a,c;\, z) = \frac{1}{\Gamma(a)} \int_0^{\infty} t^{a-1}(1+t)^{c-a-1} \exp(-zt) \, dt \tag{3.20}$$

(see Erdelyi et al (1953, p. 255) or Lebedev (1972, p. 268)).

Proof. The result is obtained by writing the joint distribution of X and Y and making the transformation

$$Z = Xt, \; t = Y/(1-Y).$$

Remark 5. The function $\Psi\,(a,c;\, z)$ exists for all a and c and has the following representations in infinite series

$$\Psi(a,c,z) = \frac{\Gamma(1-c)}{\Gamma(1+a-c)} \; {}_1F_1(a,c;\, z) + \frac{\Gamma(c-1)}{\Gamma(a)} \; z^{1-c} \; {}_1F_1(1+a-c,2-c;\, z)$$

provided $c \neq 0, \pm 1, \pm 2, \ldots$ and $\Gamma(c+1) = c\Gamma(c)$ for any $c \neq 0, \pm 1, \ldots$.

$$\Psi(a,n+1,z) = \frac{(-1)^n}{\Gamma(a-n)} \sum_{k=0}^{\infty} \frac{(a)_k z^k}{k!(n+x)!} \, [\gamma(a+k) - \gamma(1+k) - \gamma(n+1+k) + \log z]$$

$$+ \frac{1}{\Gamma(a)} \sum_{k=0}^{n-1} \frac{(-1)^k (n-k-1)!(a-n)_k}{k!} \, z^{k-n}$$

if $n = 0,1,2,\ldots$ and $a \neq 0,-1,-2,\ldots$, where $\gamma(x) = \Gamma'(x)/\Gamma(x)$, and the last term is zero if $n = 0$.

If a = -m, m = 0,1,..., and c = n + 1, n = 0,1,..., then

$$\Psi(-m,n+1;z) = (-1)^m \frac{(m+n)!}{n!}\ {}_1F_1(-m,n+1;z)$$

where

$${}_1F_1(a,c;z) = \frac{\Gamma(c)}{\Gamma(a)\Gamma(c-a)} \int_0^1 e^{zt} t^{a-1}(1-t)^{c-a-1} dt.$$

4. TESTS FOR ADDITIONAL INFORMATION

Let us consider the case of discrimination of a given signal from pure noise. A question of some practical importance is the number of features to be measured. Let us consider a signal δ with $p = r + s$ features and an estimate $f^{-1}S$ of the unknown Σ based on f degrees of freedom (or f samples from noise process) in partitioned forms

$$\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}, \quad S = \begin{pmatrix} S_{11} & S_{12} \\ S_{21} & S_{22} \end{pmatrix} \tag{4.1}$$

where δ_1 is an r-vector, δ_2 is an s-vector, Σ_{11} is an r x r matrix and so on. The signal to noise ratio based on δ (all the features) is $\delta'\Sigma^{-1}\delta$ while that based on δ_2 is $\delta_2'\Sigma_{22}^{-1}\delta_2$. If δ_1 is redundant, then

$$0 = \delta'\Sigma^{-1}\delta - \delta_2'\Sigma_{22}^{-1}\delta_2$$

$$= (\delta_1-\beta\delta_2)'\Sigma_{2\cdot1}^{-1}(\delta_1-\beta\delta_2), \quad \beta = \Sigma_{12}\Sigma_{22}^{-1} \tag{4.2}$$

which implies that $\delta_1 = \beta\delta_2$. We develop a test of the null hypothesis

$$H_0\colon\ \delta_1 = \beta\delta_2 \tag{4.3}$$

on the basis of the information provided by S.

We first consider the case where δ and S are real. From Lemma 1, $S_{1\cdot 2}$ and $b = S_{12}S_{22}^{-1}$ are independently distributed for given S_{22} with

$$S_{1\cdot 2} \sim W_r(f-s,\Sigma_{1\cdot 2}),$$

$$b \sim N_{r,s}(\beta,\Sigma_{1\cdot 2},S_{22}^{-1}). \tag{4.4}$$

Then from the standard MANOVA theory (see Rao (1973, pp. 547-550) and Srivastava and Khatri (1979, pp. 166-172)), the test statistic for testing H_0 in (4.3) is

$$T^2 = \frac{f-p+1}{r}\,\frac{(\delta_1-b\delta_2)'S_{1\cdot 2}^{-1}(\delta_1-b\delta_2)}{\delta_2'S_{22}^{-1}\delta_2} \tag{4.5}$$

which has Hotelling's T^2 or F distribution on r and (f-p+1) degrees of freedom. An alternative way of computing (4.5) is

$$T^2 = \frac{f-p+1}{r}\left[\frac{\delta'S^{-1}\delta}{\delta_2'S_{22}^{-1}\delta_2}-1\right]. \tag{4.6}$$

The test (4.5) is important since in practical applications with an estimated covariance matrix, inclusion of too many features may reduce the power of discrimination (see Rao (1971)).

Let us consider the case of k signals represented by the columns of a p x k matrix Δ. Writing

$$\Delta = \begin{pmatrix}\Delta_1\\ \Delta_2\end{pmatrix} \tag{4.7}$$

where Δ_1 is r x k matrix, we ask the question whether Δ_1 is redundant. The test for this again follows from the general MANOVA theory (see Rao (1973, pp. 547-550) and Srivastava and Khatri (1979, pp. 166-172)). The likelihood ratio test gives the Λ criterion

$$\Lambda = \frac{|S_{1\cdot 2}|}{|S_{1\cdot 2}+(\Delta_1-b\Delta_2)(\Delta_2'S_2^{-1}\Delta_2)^{-1}(\Delta_1-b\Delta_2)'|}$$

$$= \frac{|S|}{|S+\Delta\Delta'|} \div \frac{|S_{22}|}{|S_{22}+\Delta_2\Delta_2'|} \tag{4.8}$$

which is distributed as

$$\Lambda(r,f-s,k). \tag{4.9}$$

Several approximations for computing the significance of an observed value of Λ are described in Rao (1973, pp. 555-556) and Srivastava and Khatri (1979, pp. 176-186).

Remark 6. When S has complex Wishart distribution, the corresponding test for H_0: $\delta^*\Sigma^{-1}\delta = \delta_2^*\Sigma_{22}^{-1}\delta_2$ is

$$T^2 = \frac{f-p+1}{r}\,\frac{(\delta_1-b\delta_2)^*S_{1\cdot 2}^{-1}(\delta_1-b\delta_2)}{\delta_2^*S_{22}^{-1}\delta_2} \tag{4.10}$$

which has complex Hotelling's T^2 or F-distribution with 2r and 2(f-p+1) degrees of freedom. An alternative way of computing (4.10) is

$$T^2 = \frac{f-p+1}{r}\left[\frac{\delta^*S^{-1}\delta}{\delta_2^*S_{22}^{-1}\delta_2} - 1\right].$$

For the case of k signals represented by the columns of a p x k matrix $\Delta = \begin{pmatrix}\Delta_1\\ \Delta_2\end{pmatrix}$, the likelihood ratio test for $H_0:\Delta_2^*\Sigma_{22}^{-1}\Delta_2$ is

$$\Lambda = |S_{1.2}|/|S_{1.2}+(\Delta_1-b\Delta_2)(\Delta_2^* S_2^{-1}\Delta_2)^{-1}(\Delta_1-b\Delta_2)^*|$$

$$= (|S|/|S+\Delta\Delta^*|) \div (|S_{22}|/|S_{22}+\Delta_2\Delta_2^*|). \tag{4.11}$$

which is distributed as

$$\Lambda(2r,\ 2(f-r),\ k).$$

5. LOSS DUE TO ESTIMATION OF Σ IN DETECTING A SIGNAL

If Σ, the noise covariance matrix, is known, then the optimum filter for the detection of a signal δ is $\delta'\Sigma^{-1}X$ (or $\delta^*\Sigma^{-1}X$) when X is a real (or a complex) vector observation. [In the sequel we consider both the real and complex cases indicating the expressions for the complex case within brackets as above]. The signal to noise ratio, which is an index of the efficiency of discrimination, in such a case is $\delta'\Sigma^{-1}\delta$ (or $\delta^*\Sigma^{-1}\delta$). If Σ is not known but an estimate $f^{-1}S$ based on f degrees of freedom is available, we may use the estimated filter $f\delta'S^{-1}X$ (or $f\delta^*S^{-1}X$). The signal to noise ratio in such a case is

$$\rho(S,\Sigma) = \frac{(\delta' S^{-1}\delta)^2}{\delta' S^{-1}\Sigma S^{-1}\delta}, \quad \left(\text{or } \tilde{\rho}(S,\Sigma) = \frac{(\delta^* S^{-1}\delta)^2}{\delta^* S^{-1}\Sigma S^{-1}\delta}\right). \tag{5.1}$$

By the Cauchy-Schwartz inequality this is less than

$$\delta'\Sigma^{-1}\delta \ (\text{or } \delta^*\Sigma^{-1}\delta) \tag{5.2}$$

so that there is loss of information in using $f^{-1}S$ instead of Σ.

The efficiency of the estimated filter can be examined by considering the ratio of (5.1) to (5.2)

$$B = \frac{(\delta' S^{-1}\delta)^2}{(\delta'\Sigma^{-1}\delta)(\delta'\delta^{-1}\Sigma S^{-1}\delta)}, \left(\text{or } \frac{(\delta^* S^{-1}\delta)^2}{(\delta^*\Sigma^{-1}\delta)(\delta^* S^{-1}\Sigma S^{-1}\delta)} \right) \tag{5.3}$$

Using Theorem 1, (3.4), by putting $r = 1$ and $s = p-1$, the distribution of (5.3) is obtained as univariate beta

$$B_1\left(\frac{f-p+2}{2}, \frac{p-1}{2}\right), \text{ (or } B_1(f-p+2,\ p-1)). \tag{5.4}$$

The distribution (5.4) in the complex case was earlier obtained by Reed, Mallett and Brennan (1974). By computing the expected value of the distribution, they provided the rule $f \doteqdot 2p$ for maintaining an average loss ratio of better than half. But the distributions (5.4) can be used in other ways. For instance, by using incomplete beta tables one can determine the value of f, the number of samples on noise for estimating Σ, to ensure for any given p an efficiency larger than any given value with an assigned probability.

The signal to noise ratio (5.1) for any realized value S depends on the unknown Σ, which makes it difficult to assess the performance of any particular estimated filter. We suggest two ways of drawing inference on (5.1) in terms of known quantities.

First, we may find a constant c such that

$$E\,[\rho(S,\Sigma) - cf\delta' S^{-1}\delta]^2, \text{ (or } E[\tilde{\rho}(S,\Sigma) - \tilde{c}f\delta^* S^{-1}\delta]^2) \tag{5.5}$$

is a minimum. The optimum c is

$$\frac{E\,[\rho(S,\Sigma)\cdot\delta' S^{-1}\delta]}{fE(\delta' S^{-1}\delta)^2}, \left(\text{or } \frac{E\,[\tilde{\rho}(S,\Sigma)\cdot\delta^* S^{-1}\delta]}{fE(\delta^* S^{-1}\delta)^2}\right) \tag{5.6}$$

which is easily evaluated using the independence of $\rho(S,\Sigma)$ and $\delta' S^{-1}\delta$ (or $\tilde{\rho}(S,\Sigma)$ and $\delta^* S^{-1}\delta$) and the distributions derived in Theorem 1, (3.3) and (3.4) or (3.8) and (3.9), by choosing $r = 1$ and $s = p-1$. The value of c turns out to be

$$\frac{(f-p+2)(f-p-3)}{f(f+1)} \tag{5.7}$$

in either case. Then defining the estimated Mahalanobis distance $D_p^2 = f\delta' S^{-1}\delta$ (or $f\delta^* S^{-1}\delta$), we can use the known quantity

$$\frac{(f-p+2)(f-p-3)}{f(f+1)} \; D_p^2 = (1 - \frac{p-1}{f+1})(1 - \frac{p+3}{f})D_p^2 \tag{5.8}$$

as an approximation to $\rho(S,\Sigma)$ (or $\tilde{\rho}(S,\Sigma)$) for judging the efficiency of an estimated filter. Note that if f is not large compared to p, then D_p^2 overestimates the efficiency of discrimination.

The formula (5.8) is also useful in examining the gain in discrimination efficiency by increasing the number of features. For instance, the estimated signal to noise ratio with a subset of r features out of p, represented by a vector δ_1 is

$$\frac{(f-r+2)(f-r-3)}{f(f+1)} \; D_r^2 \tag{5.9}$$

where $D_r^2 = f\delta_1' S_{11}^{-1}\delta_1$ (or $f\delta_1^* S_{11}^{-1}\delta_1$) with S_{11} as the partition of S arising out of the first r columns and r rows. If $p > r$, then $D_p^2 \geq D_r^2$ but (5.9) may be > or < or = (5.8), and an appropriate decision may be taken depending on the actual relationship. It is possible that with an estimated S, the inclusion of a large number of features may decrease the discrimination efficiency, a phenomenon observed in several multivariate situations (see Rao (1971)).

A more satisfactory approach is to determine a confidence bound for $\rho(S,\Sigma)$, (or $\tilde{\rho}(S,\Sigma)$) in terms of known quantities. This is done by using the distribution derived in Theorem 3 of Section 2.

From (5.4)

$$Y = \frac{\rho(S,\Sigma)}{\delta'\Sigma^{-1}\delta} \sim B_1(\frac{f-p+2}{2}, \frac{p-1}{2}), \left(\text{or } \tilde{Y} = \frac{\tilde{\rho}(S,\Sigma)}{\delta^*\Sigma^{-1}\delta} \sim B_1(f-p+2, p-1)\right) \tag{5.10}$$

as beta variables, and

$$X = \frac{\delta'\Sigma^{-1}\delta}{\delta' S^{-1}\delta} \sim G(\tfrac{1}{2}, \tfrac{f-p+1}{2}), \left(\text{or } \tilde{X} = \frac{\delta^{*}\Sigma^{-1}\delta}{\delta^{*} S^{-1}\delta} \sim G(1, f-p+1)\right) \tag{5.11}$$

as gamma variables using the notation of Rao (1973, p. 164), and, further, X and Y (or $\tilde{X}$ and $\tilde{Y}$) are independent. Then from Theorem 3

$$Z = \frac{1}{2} XY = \frac{f}{2} \frac{\rho(S,\Sigma)}{D_p^2} \text{ or } \tilde{Z} = \tilde{X}\tilde{Y} = \frac{f\rho(S,\Sigma)}{D_p^2} \tag{5.12}$$

where $D^2 = f\delta' S^{-1}\delta$ (or $f\delta^{*} S^{-1}\delta$), has the confluent hypergeometric distribution (3.19)

$$\frac{1}{\Gamma(m)} e^{-z} z^{m-1} \frac{\Gamma(a+m-c+1)}{\Gamma(m-c+1)} \Psi\,(a,c;\, z) \tag{5.13}$$

which is independent of the unknown Σ with

$$m = \frac{f-p+1}{2},\ a = \frac{p-1}{2},\ c = \frac{1}{2},\ (\text{or } m = f-p+1,\ a = p-1,\ c = 0). \tag{5.14}$$

If z_α (or $\bar{z}_\alpha$) is the lower α % point of the distribution, then

$$P(\rho(S,\Sigma) \geq \frac{2z_\alpha}{f} D_p^2) = 1-\alpha, \left(\text{or } P(\tilde{\rho}(S,\Sigma) \geq \frac{\tilde{z}_\alpha}{f} D_p^2) = 1-\alpha\right) \tag{5.15}$$

so that

$$\rho(S,\Sigma) \geq \frac{2z_\alpha}{f} D_p^2, \left(\text{or } \tilde{\rho}(S,\Sigma) \geq \frac{\tilde{z}_\alpha}{f} D_p^2\right) \tag{5.16}$$

provides a lower bound to the realized signal to noise ratio at a confidence level of $(1-\alpha)$%.

The equation satisfied by z_α is

$$\alpha = \int_0^1 \frac{\Gamma(\frac{f+1}{2})}{\Gamma(\frac{p-1}{2})\Gamma(\frac{f-p+2}{2})} \quad y^{(f-p)/2}(1-y)^{(p-3)/2}dy \int_0^{z_\alpha/y} \frac{1}{\Gamma(\frac{f-p+1}{2})} e^{-x} x^{(f-p-1)/2} dx$$

and $\tilde{z}_\alpha$ is

$$\int_0^1 \frac{\Gamma(f+1)}{\Gamma(p-1)\Gamma(f-p+2)} y^{f-p+1}(1-y)^{p-2}dy \int_0^{\tilde{z}_\alpha/y} \frac{1}{\Gamma(f-p+1)} e^{-x} x^{f-p} dx.$$

The values of z_α (or $\tilde{z}_\alpha$) can be found by a suitable computer algorithm. For instance, the multiplying coefficients (see 5.16) for the observed Mahalanobis distance to provide 50% and 95% lower confidence bounds to the realized signal to noise ratio are given below for $p = 4$ and $f = 8$, 12 and 16.

	$2z_\alpha/f$		$\tilde{z}_\alpha/f$	
f	50%	95%	50%	95%
8	.345	.075	.381	.141
12	.525	.188	.553	.283
16	.631	.281	.649	.377

Detailed tables will appear in a later communication.

6. LOSS DUE TO ESTIMATION OF Σ IN MULTIPLE DISCRIMINATION

Consider the problem of identifying a received message as noise or one of r possible signals $\delta_1, \ldots, \delta_r$ which we represent by a $p \times r$ matrix $\Delta = (\delta_1 : \ldots : \delta_r)$. Further, let X be a vector of observed features with covariance matrix Σ and $E(X) = \delta_i$ when the i-th signal is transmitted, $i = 1, \ldots, r$ and $E(X) = 0$ for noise. Then the overall efficiency of discrimination using X can be judged by a function of the eigen values of

$$\Delta\Delta' \text{ (or } \Delta\Delta^*\text{) with respect to } \Sigma \tag{6.1}$$

which are the same as the eigen values of

$$\Delta'\Sigma^{-1}\Delta \text{ (or } \Delta^*\Sigma^{-1}\Delta\text{)}. \tag{6.2}$$

This provides a generalization of the signal to noise ratio $\delta'\Sigma^{-1}\delta$ (or $\delta^*\Sigma^{-1}\delta$) in the case of a single signal.

If the noise has $N_p(0,\Sigma)$ distribution, then the decision function for the detection of r signals is based on the sufficient statistics

$$\delta_i'\Sigma^{-1}X \text{ (or } \delta_i^*\Sigma^{-1}X\text{)}, \; i = 1,\dots,r \tag{6.3}$$

which can be written as the discriminant vector $Y = \Delta'\Sigma^{-1}X$ (or $\Delta^*\Sigma^{-1}X$) with the covariance matrix $\Delta'\Sigma^{-1}\Delta$ (or $\Delta^*\Sigma^{-1}\Delta$), and $E(Y) = \Delta'\Sigma^{-1}\delta_i$ (or $\Delta^*\Sigma^{-1}\delta_i$) for the i-th signal. The efficiency of discrimination in using Y instead of X, using the formula (6.1) depends on the eigen values of

$$(\Delta'\Sigma^{-1}\Delta)(\Delta'\Sigma^{-1}\Delta)^{-1}(\Delta'\Sigma^{-1}\Delta) \text{ with respect } \Delta'\Sigma^{-1}\Delta \tag{6.4}$$

(or with Δ^* in the place of Δ), which are the same as those for X as expected. If Σ is not known but an estimate $f^{-1}S$ is available, then the estimated discriminant vector is

$$\hat{Y} = \Delta'S^{-1}X \text{ (or } \Delta^*S^{-1}X\text{)} \tag{6.5}$$

and its efficiency depends on the eigen values of

$$B = (\Delta'S^{-1}\Delta)(\Delta'S^{-1}\Sigma S^{-1}\Delta)^{-1}\Delta'S^{-1}\Delta \tag{6.6}$$

(or with S' replaced by S^*), which is a generalization of $\rho(S,\Sigma)$, (or $\tilde{\rho}(S,\Sigma)$) as considered in (5.1).

In Theorem 1, we found the distribution of the matrices $(\Delta' S^{-1}\Delta)^{-1}$ and $(\Delta'\Sigma^{-1}\Delta)^{\frac{1}{2}} B (\Delta'\Sigma^{-1}\Delta)^{\frac{1}{2}}$ in the case of real variables, and of the matrices $(\Delta^* S^{-1}\Delta)^{-1}$ and $(\Delta^*\Sigma^{-1}\Delta)^{\frac{1}{2}} B (\Delta^*\Sigma^{-1}\Delta)^{\frac{1}{2}}$ in the complex case. We use these distributions in examining the realized efficiency through the estimated discriminant vector.

For this purpose, we consider two particular functions of the eigen values of B, one of which is the sum

$$Z_1 = \operatorname{tr} B = \operatorname{tr} [(\Delta' S^{-1}\Delta)^2 (\Delta' S^{-1}\Sigma S^{-1}\Delta)^{-1}]$$

$$= \sum_1 \delta_i' S^{-1}\Delta(\Delta' S^{-1}\Sigma S^{-1}\Delta)^{-1}\Delta' S^{-1}\delta_i \qquad (6.7)$$

(or with δ_i' and Δ' replaced by δ_i^* and Δ^*), and another is the product

$$Z_2 = |B| = \frac{|\Delta' S^{-1}\Delta|^2}{|\Delta' S^{-1}\Sigma S^{-1}\Delta|}, \left(\text{or } \frac{|\Delta^* S^{-1}\Delta|}{|\Delta^* S^{-1}\Sigma S^{-1}\Delta|}\right). \qquad (6.8)$$

Using Theorem 2

$$E(Z_1) = \frac{f-p+2r}{f+r} \left[\sum_1^r \delta_i'\Sigma^{-1}\delta_i \ (\text{or } \delta_i^*\Sigma^{-1}\delta_i)\right]$$

$$= \frac{f-p+2r}{f+r} [\operatorname{tr}(\Delta'\Sigma^{-1}\Delta)(\text{or } \Delta^*\Sigma^{-1}\Delta)] \qquad (6.9)$$

and

$$E(Z_2) = \left[\prod_{i=1}^r \frac{f-p+2r-i+1}{f+r-i+1}\right]\left(|\Delta'\Sigma^{-1}\Delta|^2 \text{ or } |\Delta^*\Sigma^{-1}\Delta|^2\right). \qquad (6.10)$$

The formulas (6.9) and (6.10) enable us to choose a suitable value of f for given p and r to keep the average loss at a desired level.

7. REFERENCES

Erdélyi, A., Magnus, W., Oberhettinger, F., and Tricomi, F.G. (1953). Higher Transcendental Functions, Vol. I. McGraw-Hill book company, Inc., New York.

Khatri, C.G. and Pillai, K.C.S. (1965). Some results of the noncentral multivariate beta distribution and moments of traces of two matrices. Annals of Math. Statist. 36, 1511-1520.

Lebedev, N.N. (1972). Special Functions and Their Applications. Dover Publications, New York. (Translated and edited by Richard A. Silverman).

Rao, C. Radhakrishna (1971). Advanced Statistical Methods in Biometric Research. Heffner, New York.

Rao, C. Radhakrishna (1973). Linear Statistical Inference and its Applications (Second Edition). Wiley, New York.

Reed, I.S., Mallett, J.D. and Brennan, L.E. (1974). Rapid convergence rate in adaptive rays. IEEE Transactions on Aerospace and Electronic Systems. 10, 853-863.

Srivastava, M.S. and Khatri, C.G. (1979). Introduction to Multivariate Statistics. North Holland Publishing Company, New York.

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE

READ INSTRUCTIONS BEFORE COMPLETING FORM

1. REPORT NUMBER: [redacted]

2. GOVT ACCESSION NO.:

3. RECIPIENT'S CATALOG NUMBER:

4. TITLE (and Subtitle): Effects of Estimated Noise Covariance Matrix in Optimal Signal Detection

5. TYPE OF REPORT & PERIOD COVERED: Technical - October, 1985

6. PERFORMING ORG. REPORT NUMBER: 85-38

7. AUTHOR(s): C.G. Khatri and C. Radhakrishna Rao

8. CONTRACT OR GRANT NUMBER(s): [redacted] F49620-85-C-0008 [redacted]

9. PERFORMING ORGANIZATION NAME AND ADDRESS:
Center for Multivariate Analysis
515 Thackeray Hall
Univ. of Pittsburgh, Pittsburgh, PA 15260

10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS:

11. CONTROLLING OFFICE NAME AND ADDRESS:
[redacted]
Air Force Office of Scientific Research
Bo AFB DC 20332

12. REPORT DATE: October, 1985

13. NUMBER OF PAGES: 25

14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office):

15. SECURITY CLASS. (of this report): Unclassified

15a. DECLASSIFICATION/DOWNGRADING SCHEDULE:

16. DISTRIBUTION STATEMENT (of this Report):

Approved for public release; distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report):

18. SUPPLEMENTARY NOTES:

19. KEY WORDS (Continue on reverse side if necessary and identify by block number):

20. ABSTRACT (Continue on reverse side if necessary and identify by block number):

There is loss of efficiency when an estimated noise covariance matrix is used in the place of the unknown true noise covariance matrix in the construction of the optimum filter for signal detection. In the case of detecting a single signal specified by a real or a complex vector, we investigate the extent of this loss by obtaining an exact confidence bound for the realized signal to noise ratio. We also give an estimate of this ratio which is useful in optimum selection of features. Some of these results are extended to the case of discrimination between a number of given signals.

DD FORM 1 JAN 73 1473

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

END

FILMED

7-86

DTIC